

Information channel capacity in the field theory estimation

J. Śladkowski ^{a,*}, J. Syska ^{b,**}

^a*Department of Astrophysics and Cosmology, Institute of Physics, University of Silesia, Uniwersytecka 4, 40-007 Katowice, Poland*

^b*Department of Field Theory and Particle Physics, Institute of Physics, University of Silesia, Uniwersytecka 4, 40-007 Katowice, Poland*

Abstract

The construction of the information capacity for the vector position parameter in the Minkowskian space-time is presented. This lays the statistical foundations of the kinematical term of the Lagrangian of the physical action for many field theory models, derived by the extremal physical information method of Frieden and Soffer.

Keywords: channel information capacity, Fisher information, Stam's information, causality

1. Introduction

The Fisher information (I_F) is a second degree covariant tensor, which is one of the contrast functions defined on the statistical space \mathcal{S} [1]. It is the local version of the Kulback-Leibler entropy for two distributions infinitesimally displaced from one another [2, 3]. This, in turn, is the model selection tool used in the standard maximum likelihood (ML) method and the basic notion in the definition of the information channel capacity [3, 4].

The method of nonparametric estimation that enables the statistical selection of the equation of motions (or generating equations) of various field theory or statistical physics models is called the extremal physical information (EPI). The central quantity of EPI analysis is the information channel capacity I , which is the trace of the expectation value of the I_F matrix. Fundamentally, it enters into the EPI formalism as the second order coefficient in the Taylor expansion of the log-likelihood function [5]. Originally, EPI was proposed by Frieden and Soffer [2]. They used two Fisherian information quantities: the intrinsic information J of the source phenomenon and the information channel capacity I , which connects the phenomenon and observer. Both J and I , together with their densities, are used in the construction of two information principles, the

*Corresponding author: J. Śladkowski (*Email address:* jan.sladkowski@us.edu.pl)

**Corresponding author: J. Syska (*Email address:* jacek.syska@us.edu.pl)

structural and variational ones. J and I are the counterparts¹ of the Boltzmann and Shannon entropies, respectively, however, they are physically different from them [2, 3]. Finally, although in the Frieden and Soffer approach the structural information principle is postulated on the expected level, it is then reformulated to the observed one, together with the variational principle giving two coupled differential equations.

In [7] two information principles were also postulated, but with a different interpretation of the *structural information*, which is denoted by Q . In [5] the derivations, firstly of the observed and secondly of the expected *structural information principle* from basic principles was given. It was based on the analyticity of the logarithm of the likelihood function, which allows for its Taylor expansion in the neighborhood of the true value of the vector parameter Θ , and on the metricity of the statistical space \mathcal{S} of the system. The analytical structure of the structural information principle and the geometrical meaning of the variational information principle, which leads to the Euler-Lagrange equations, are discussed in [8, 9]. Both information principles made the EPI method the fundamental tool in the physical model selection, which is a kind of non-parametric estimation, having, as the output of the solution of the information principles the equations of motion or distribution generating equation. Their usage for the derivation of the basic field theory equations of motion is thus a fundamental one, as they anticipate these equations [2, 8, 10]. The fact that the formalism of the information principles is used for the derivation of the distribution generating equation signifies that² the microcanonical description of the thermodynamic properties of a compound system has to meet the analyticity and metricity assumptions as well.

Thus, it is obvious from the former work that both the form of I and its density [5, 8, 9] play crucial roles in the construction of a particular physical model. Frieden, Soffer along with Plastino and Plastino [2] put into practice the solution of the differential information principles for various EPI models. Previously, in [2] the role of I was discussed in many field theory contexts and in [1, 5, 8] the general view on the construction of I was also presented. The main topic of the present paper will concentrate on the construction of I for field theory models with the estimation performed in the base space with the Minkowski metric.

An important model parameter in the EPI analysis is the dimension N of the sample which via the likelihood function of the system enters into the channel information capacity I . The physical models form two separate categories with respect to N . In the first one, to which wave mechanics and field theories belong, both N and I are finite [8]. Classical mechanics, on the base space \mathcal{Y} continuum, forms the second class³ having infinite N and thus infinite I . This

¹In the sense that they are similar in relating them [6].

²In agreement with the Jaynes' principle [6].

³In the contrast to e.g. wave mechanics, in classical mechanics the solution of the equation of motion does not determine (from this equation) the structure of a particle, which has to be determined independently at every point of the particle trajectory by the definition of its point structure, e.g. by the means of the δ -Dirac distribution.

fact was applied to prove the impossibility of the derivation of wave and field theory models from classical mechanics in [7]. In the case of the first category, the sample size N is the rank of the field of the system [2]. For example, with $N = 8$, the EPI method can result in the Dirac wave equation, whereas with $N = 32$ it can result in the Rarita-Schwinger one [2]. In the realm of statistical physics, e.g. for $N = 1$, the equilibrium Maxwell-Boltzmann velocity law is obtained, while for $N > 1$, the non-equilibrium, although still stationary solutions (that otherwise follow from the Boltzmann transport equation) were discovered [2]. Since the observed structural information was obtained and the new interpretation of the structural information Q established [5, 9, 8], some of these models have been recalculated [8, 10, 11]. It appears that for every field, N can be related to the dimension of the representation of the group of symmetry transformation of the field in question [2].

The paper will also deal with the kinematic form of channel information capacity I expressed in terms of the point probability distributions of the sample. For $N = 1$, the usefulness of this form is perceived in the proof of the **I**-theorem [2, 8], which is the informational analog of the **H**-Boltzmann theorem. Then, the Fisher temperature, which is the sensitivity of the Fisher information to the change of the scalar parameter, can be defined. Thus, the **I**-theorem is in a sense more general than its older thermodynamic predecessor, as it can describe not only the thermodynamic properties of the compound system [12] but also of an elementary one-particle system.

2. The basics: Rao-Fisher metric and Rao-Cramér theorem

Suppose that the original random variable Y takes vector values $\mathbf{y} \in \mathcal{Y}$ and let the vector parameter θ of the distribution $p(\mathbf{y})$, in which we are interested be the *expected parameter*, i.e. the expectation value of Y :

$$\theta \equiv E(Y) = \int_{\mathcal{Y}} d\mathbf{y} p(\mathbf{y}) \mathbf{y} . \quad (1)$$

Let us now consider the N -dimensional sample $\tilde{Y} = (Y_1, Y_2, \dots, Y_N) \equiv (Y_n)_{n=1}^N$, where every Y_n is the variable Y in the n -th population, $n = 1, 2, \dots, N$, which is characterized by the value of the vector parameter θ_n . The specific realization of \tilde{Y} takes the form $y = (\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_N) \equiv (\mathbf{y}_n)_{n=1}^N$, where every datum \mathbf{y}_n is generated from the distribution $p_n(\mathbf{y}_n|\Theta)$ of the random variable Y_n , where the vector parameter Θ is given by:

$$\begin{aligned} \Theta &= (\theta_1, \theta_2, \dots, \theta_N)^T \equiv (\theta_n)_{n=1}^N , \\ \theta_n &= (\vartheta_{1n}, \vartheta_{2n}, \dots, \vartheta_{kn})^T \equiv ((\vartheta_s)_{s=1}^k)_n . \end{aligned} \quad (2)$$

The set of all possible realizations y of the sample \tilde{Y} forms the sample space \mathcal{B} of the system. In this paper, we assume that the variables Y_n of the sample \tilde{Y} are independent. Hence, the expected parameter $\theta_{n'} = \int_{\mathcal{B}} dy P(y|\Theta) \mathbf{y}_{n'}$ does not influence the *point probability* distribution $p_n(\mathbf{y}_n|\theta_n)$ for the sample

index $n' \neq n$. The data are generated in agreement with the point probability distributions, which fulfill the condition:

$$p_n(\mathbf{y}_n|\Theta) = p_n(\mathbf{y}_n|\theta_n), \quad \text{where } n = 1, \dots, N, \quad (3)$$

and the likelihood function $P(y|\Theta)$ of the sample $y = (\mathbf{y}_n)_{n=1}^N$ is the product:

$$P(\Theta) \equiv P(y|\Theta) = \prod_{n=1}^N p_n(\mathbf{y}_n|\theta_n). \quad (4)$$

The set of values of $\Theta = (\theta_n)_{n=1}^N$ forms the coordinates of $P(y|\Theta)$, which is a point in $d = k \times N$ - dimensional statistical (sub)space \mathcal{S} [1]. The number of all parameters is equal to $d = k \times N$, but for simplicity, we will use the notation $\Theta \equiv (\theta_1, \theta_2, \dots, \theta_d)^T \equiv (\theta_i)_{i=1}^d$, where the index $i = 1, 2, \dots, d$ replaces the pair of indexes "sn". The likelihood function is formally the joint probability distribution of the realization $y \equiv (\mathbf{y}_n)_{n=1}^N$ of the sample $\tilde{Y} \equiv (Y_n)_{n=1}^N$; hence, P is the probability measure on \mathcal{B} . The set of all measures $\Sigma(\mathcal{B})$ on \mathcal{B} is the state space of the model.

The Fisher information matrix: Let us now examine a subset $\mathcal{S} \subset \Sigma(\mathcal{B})$, on which the coordinate system $(\xi^i)_{i=1}^d$ is given [1] so that the statistical space \mathcal{S} is the d - dimensional manifold⁴. Assume that \mathcal{B} the d - dimensional statistical model:

$$\mathcal{S} = \{P_\Theta \equiv P(y|\Theta), \Theta \equiv (\theta_i)_{i=1}^d \in V_\Theta \subset \mathbb{R}^d\}, \quad (5)$$

is given, i.e. the family of the probability distributions parameterized by d non-random variables $(\theta_i)_{i=1}^d$ which are real-valued and belong to the parametric space V_Θ of the parameter Θ , i.e. $\Theta \in V_\Theta \subset \mathbb{R}^d$. Thus, the logarithm of the likelihood function $\ln P : V_\Theta \rightarrow \mathbb{R}$ is defined on the space V_Θ .

Let $\tilde{\Theta} \equiv (\tilde{\theta}_i)_{i=1}^d \in V_\Theta$ be another value of the parameter or a value of the estimator $\tilde{\Theta}$ of the parameter $\Theta = (\theta_i)_{i=1}^d$. At every point, P_Θ , the $d \times d$ - dimensional observed Fisher information (FI) matrix can be defined [13, 5, 8]:

$$\mathfrak{IF}(\Theta) \equiv -\partial^{i'} \partial^i \ln P(\Theta) = \left(-\tilde{\partial}^{i'} \tilde{\partial}^i \ln P(\tilde{\Theta}) \right)_{\big|_{\tilde{\Theta}=\Theta}} \quad (6)$$

and $\partial^i \equiv \partial/\partial\theta_i$, $\tilde{\partial}^i \equiv \partial/\partial\tilde{\theta}_i$, $i, i' = 1, 2, \dots, d$. It characterizes the local properties of $P(y|\Theta)$. It is symmetric and in field theory and statistical physics models with continuous, regular [13] and normalized distributions, it is positively definite. We restrict the considerations to this case only. The expected $d \times d$ - dimensional FI matrix on \mathcal{S} at point P_Θ [1] is defined as follows:

$$I_F(\Theta) \equiv E_\Theta(\mathfrak{IF}(\Theta)) = \int_{\mathcal{B}} dy P(y|\Theta) \mathfrak{IF}(\Theta), \quad (7)$$

⁴ In this paper, we are interested in only the global coordinate systems.

where the differential element $dy \equiv d^N \mathbf{y} = d\mathbf{y}_1 d\mathbf{y}_2 \dots d\mathbf{y}_N$. The subscript Θ in the expected value signifies the true value of the parameter under which the data y are generated. The FI matrix defines on \mathcal{S} the riemannian Rao-Fisher metric g^{ij} , which in the coordinates $(\theta_i)_{i=1}^d$ has the form $(g^{ij}(\Theta)) := I_F$. Under regularity and normalization conditions $\int_{\mathcal{B}} d^N \mathbf{y} P(\Theta) = 1$, [13], we obtain $\int_{\mathcal{B}} dy P(\Theta) \partial^i \ln P(\Theta) = 0$, $i = 1, 2, \dots, d$. Therefore, the elements g^{ij} of I_F can be rewritten as follows:

$$\begin{aligned} g^{ij}(\Theta) &= -E_{\Theta} (\partial^i \partial^j \ln P(y|\Theta)) = - \int_{\mathcal{B}} dy P(y|\Theta) \partial^i \partial^j \ln P(y|\Theta) \quad (8) \\ &= \int_{\mathcal{B}} dy P(y|\Theta) \partial^i \ln P(y|\Theta) \partial^j \ln P(y|\Theta) \\ &= E_{\Theta} (\partial^i \ln P(y|\Theta) \partial^j \ln P(y|\Theta)) \quad , \quad i, j = 1, 2, \dots, d \quad , \quad \forall P_{\Theta} \in \mathcal{S} . \end{aligned}$$

Owing to the last line, $\mathbf{iF} = (\mathbf{iF}^{i'i})$ is sometimes recorded in the "quadratic" form:

$$\mathbf{iF} = \left(\partial^{i'} \ln P(\Theta) \partial^i \ln P(\Theta) \right) \quad , \quad (9)$$

as it is useful in the definition of the α -connection on the statistical space \mathcal{S} [1].

The multiparametric Cramér-Rao theorem: Let $I_F(\Theta)$ be the Fisher information matrix (7) for $\Theta \equiv (\theta_i)_{i=1}^d$ and $\hat{\theta}_i = \hat{\theta}_i(\tilde{Y})$ be an unbiased estimator of the distinguished parameter θ_i :

$$E_{\Theta} \hat{\theta}_i = \theta_i \in \mathfrak{R} \quad (10)$$

and the values of the remaining parameters may also be unknown, i.e. they are to be estimated from the sample simultaneously with θ_i . Then, the variance of $\hat{\theta}_i$ fulfills the Cramér-Rao (CR) inequality $\sigma_{\Theta}^2(\hat{\theta}_i) \geq [I_F^{-1}(\Theta)]_{ii} =: I_F^{ii}(\Theta)$, where I_F^{-1} is the inverse of I_F [13]. $I_F^{ii}(\Theta)$ is the *lower bound in the Cramér-Rao (CRLB) inequality* of the variance of the estimator $\hat{\theta}_i$ [13].

Let $I_F^i \equiv I_F^{ii}(\Theta)$ and $I_{Fi} \equiv I_{Fii}(\theta_i) = I_{Fii}(\Theta)$ denote the (i, i) elements of the matrix $I_F^{-1}(\Theta)$ and $I_F(\Theta)$, respectively. In the multiparametric case discussed, the pair of inequalities proceed [13, 8]:

$$\sigma_{\Theta}^2(\hat{\theta}_i) \geq I_F^i \geq \frac{1}{I_{Fi}} \quad , \quad \text{where } 1 \leq i \leq d \quad , \quad (11)$$

where the first one is the CR inequality. If Θ is the scalar parameter, or if the θ_i parameter is estimated only and the others are known, then in the second inequality in (11) the equality $I_F^i = 1/I_{Fi}$ remains. In this paper, we maintain the name of the Fisher information of the parameter θ_i for I_{Fi} regardless of whether the other parameters are simultaneously estimated.

3. The channel information capacity for the position random variable

Let the vector value $\mathbf{y} \in \mathcal{Y}$ of Y be the space position vector. It could be the space-time point $\mathbf{y} \equiv (\mathbf{y}^\nu)_{\nu=0}^3$ of the four-dimensional Minkowski space $\mathcal{Y} \equiv \mathbb{R}^4$, which occurs in the description of the system, e.g. in wave mechanics, or the space point $\mathbf{y} \equiv (\mathbf{y}^\nu)_{\nu=1}^3 \in \mathcal{Y} \equiv \mathbb{R}^3$ in the three-dimensional Euclidean space⁵. Thus, $\mathbf{y} \equiv (\mathbf{y}^\nu)$ can possess the vector index $\nu, \mu = (0), 1, 2, 3, \dots$, where $\nu, \mu = 0, 1, 2, 3, \dots$ in the Minkowski space and $\nu, \mu = 1, 2, 3, \dots$ in the Euclidean one. In this analysis, we will use random variables with covariant and contravariant coordinates. The relation between them, both for the values of the random position vector and for the corresponding expectation values, is as follows:

$$\mathbf{y}_\nu = \sum_{\mu} \eta_{\nu\mu} \mathbf{y}^\mu, \quad \theta_\nu = \sum_{\mu} \eta_{\nu\mu} \theta^\mu, \quad (12)$$

where $(\eta_{\nu\mu})$ is the metric tensor of the space \mathcal{Y} . In the case of the vectorial Minkowski index, we take the following diagonal form of the metric tensor:

$$(\eta_{\nu\mu}) = \text{diag}(1, -1, -1, -1, \dots), \quad (13)$$

whereas for the Euclidean vectorial index, the metric tensor takes the form:

$$(\eta_{\nu\mu}) = \text{diag}(1, 1, 1, \dots). \quad (14)$$

The introduction of the metric tensor η is important from the measurement point of view. In the measurement of the chosen ν -th coordinate of the position, we are not able to exclude the displacements (fluctuations) of the values of the space-time coordinates which are orthogonal to it. This indicates that the expectation value of the ν -th coordinate of the position is not calculated in (1) from a distribution of the type $p(\mathbf{y}^\nu)$, but that it has to be calculated from the joint distribution $p(\mathbf{y})$ for all coordinates \mathbf{y}^ν . In addition, the measurement which is independent of the coordinate system is the one of the square length $\mathbf{y} \cdot \mathbf{y}$ and not of the single coordinate \mathbf{y}^ν only, where " \cdot " denotes the inner product defined by the metric tensor (13) or (14).

The fact of the statistical dependence of the spacial position variables for the different indexes ν should not be confused with the *analytical independence* which they possess. This means that the variable Y is the so-called *Fisherian variable* for which:

$$\partial \mathbf{y}^\nu / \partial \mathbf{y}^\mu = \delta_\mu^\nu. \quad (15)$$

Let the data $y \equiv (\mathbf{y}_n)_{n=1}^N$ be a realization of the N -dimensional sample \tilde{Y} for the *positions of the system*, where $\mathbf{y}_n \equiv (\mathbf{y}_n^\nu)$, $n = 1, 2, \dots, N$, denotes the n -th vectorial observation. Now, as the number of the parameters θ_n , where

⁵In the derivation of the equations generating the distribution in statistical physics, \mathbf{y} can be the value of the energy ϵ of the system [2] and then $\mathbf{y} \equiv \epsilon \in \mathcal{Y} \equiv \mathbb{R}$.

$n = 1, 2, \dots, N$, agrees with the dimension N of the sample, and as ν is the vectorial index of the coordinate \mathbf{y}_n^ν (and therefore the vector parameter has the additional vectorial index ν , too) thus, we have:

$$\Theta = (\theta_1, \theta_2, \dots, \theta_N) \text{ , where } \theta_n = (\theta_n^\nu) \text{ , } \nu = (0), 1, 2, 3, \dots \text{ ,} \quad (16)$$

where the expected parameter is the expectation value of the position of the system at the n -th measurement point of the sample:

$$\theta_n \equiv E(Y_n) = (\theta_n^\nu) \text{ , where } \theta_n^\nu = \int_{\mathcal{B}} dy P(y|\Theta) \mathbf{y}_n^\nu . \quad (17)$$

Here for $\mathbf{y} \equiv (\mathbf{y}^\nu)_{\nu=0}^3 \in \mathcal{Y} \equiv \mathbb{R}^4$, $dy := d^4\mathbf{y}_1 \dots d^4\mathbf{y}_N$ and $d^4\mathbf{y}_n = d\mathbf{y}_n^0 d\mathbf{y}_n^1 d\mathbf{y}_n^2 d\mathbf{y}_n^3$.

The statistical spaces \mathcal{S} and $\mathcal{S}_{N \times 4}$: Let the discussed space be the Minkowski space-time $\mathcal{Y} \equiv \mathbb{R}^4$. Then, every one of the distributions $p_n(\mathbf{y}_n|\theta_n)$ is the point of the statistical model $\mathcal{S} = \{p_n(\mathbf{y}_n|\theta_n)\}$, which is parameterized by the natural parameter, i.e. by the expectation value $\theta_n \equiv (\theta_n^\nu)_{\nu=0}^3 = E(Y_n)$, as in (17). Consequently, the dimension of the sample space \mathcal{B} and the dimension of the parametric space V_Θ of the vector parameter $\Theta \equiv (\theta_n^\nu)_{n=1}^N$ are equal to each other⁶ and, as the sample \tilde{Y} is $N \times 4$ -dimensional random variable, hence the set $\mathcal{S}_{N \times 4} = \{p_n(y|\Theta)\}$ is the statistical space on which the parameters $(\theta_n^\nu)_{n=1}^N$ form the $N \times 4$ -dimensional local coordinate system.

3.1. The Rao-Cramér inequality for the position random variable

According to (8), the channel information capacity in the *single n -th* (measurement) *information channel*, i.e. the *Fisher information for the parameter* θ_n is equal to:

$$\begin{aligned} I_{Fn} &\equiv I_{Fn}(\theta_n) = \int_{\mathcal{B}} dy P(y|\Theta) (\nabla_{\theta_n} \ln P(y|\Theta) \cdot \nabla_{\theta_n} \ln P(y|\Theta)) \\ &= \int_{\mathcal{B}} dy P(y|\Theta) \sum_{\nu, \mu=(0), 1, 2, \dots} \eta^{\nu\mu} \left(\frac{\partial \ln P(y|\Theta)}{\partial \theta_n^\nu} \frac{\partial \ln P(y|\Theta)}{\partial \theta_n^\mu} \right) , \end{aligned} \quad (18)$$

where the tensor $(\eta^{\nu\mu})$ is dual to $(\eta_{\nu\mu})$, i.e. $\sum_{\mu=(0), 1, 2, \dots} \eta_{\nu\mu} \eta^{\gamma\mu} = \delta_\nu^\gamma$, where δ_ν^γ is the Kronecker delta and $\nabla_{\theta_n} = \sum_{\mu=(0), 1, 2, \dots} \frac{\partial}{\partial \theta_n^\mu} d\mathbf{y}_n^\mu$.

Simultaneously, the variance of the estimator $\hat{\theta}_n(y)$ of the parameter θ_n has the form:

$$\begin{aligned} \sigma^2(\hat{\theta}_n) &= \int_{\mathcal{B}} dy P(y|\Theta) (\hat{\theta}_n(y) - \theta_n) \cdot (\hat{\theta}_n(y) - \theta_n) \\ &= \int_{\mathcal{B}} dy P(y|\Theta) \sum_{\nu, \mu=(0), 1, 2, \dots} \eta_{\nu\mu} (\hat{\theta}_n^\nu(y) - \theta_n^\nu) (\hat{\theta}_n^\mu(y) - \theta_n^\mu) . \end{aligned} \quad (19)$$

⁶Nevertheless, let us remember that, in general, the dimension of the vector of parameters $\Theta \equiv (\theta_i)_{i=1}^d$ and the sample vector $y \equiv (\mathbf{y}_n)_{n=1}^N$ can be different.

Let us now observe that for the variables of the space position, the integrals in (18) and (19) are not to be factorized with respect to the vectorial ν -th coordinate.

Now, as a consequence of (3), for every distinguished parameter θ_n , the variance $\sigma^2(\hat{\theta}_n)$ of its estimator given by (19) is connected with the Fisher information $I_{Fn} = I_{Fn}(\theta_n)$ given by (18) in the single information channel for this parameter by the inequality (11):

$$\frac{1}{\sigma^2(\hat{\theta}_n)} \leq \frac{1}{I_F^n} \leq I_{Fn} \quad \text{where } n = 1, 2, \dots, N, \quad (20)$$

where I_F^n is the CRLB for the parameter θ_n .

The Stam's information: The quantity $\frac{1}{\sigma^2(\hat{\theta}_n)}$ refers to the single n -th channel. The *Stam's information* I_S [14, 2] is obtained by summing over its index n and is equal to:

$$0 \leq I_S \equiv \sum_{n=1}^N \frac{1}{\sigma^2(\hat{\theta}_n)} =: \sum_{n=1}^N I_{Sn}, \quad (21)$$

where $\sigma^2(\hat{\theta}_n)$ is given in (19). This is the *scalar* measure of the quality of the simultaneous estimation in all information channels. *The Stam's information is by definition always nonnegative.* As $\theta_n = (\theta_n^\nu)$ is the vectorial parameter, $I_{Sn} = \frac{1}{\sigma^2(\hat{\theta}_n)}$ itself is the Stam's information of the (time-) space channels for the n -th measurement in the sample⁷.

Finally, summing the LHS and RHS of (20) over the index n and taking into account (21), we observe that I_S fulfills the inequality:

$$0 \leq I_S \equiv \sum_{n=1}^N I_{Sn} \leq \sum_{n=1}^N I_{Fn} =: I, \quad (22)$$

where I , denoted in the statistical literature by C , is the *channel information*

⁷The appearance of the Minkowski metric in the Stam's information I_{Sn} in (21) and (19) justifies the use of the error propagation law (and the calculation of the mean square of the measurable quantity) in the arbitrary Euclidean metric, that is, when in addition to the spacial indexes x_k , $k = 1, 2, 3, \dots$, the temporal index t occurs with the imaginary unit i [15, 2].

capacity [3, 4] of the system⁸:

$$I = \sum_{n=1}^N I_{Fn} . \quad (24)$$

The inequality (22) is the 'minimal', *trace* generalization of the "single-channel" CR inequality (20).

From the physical modeling point of view, the channel information capacity I is the most important statistical concept, which lays the foundation for the kinematical terms [2] of various field theory models. According to (21), it appears that both for the Euclidean (14) and Minkowskian metric (13) we perform the estimation in the case of positive I_S only. Hence, from (22) it follows that I is also non-negative. In (52) Section 5, it will be shown that I is non-negatively defined for the field theory with particles which have a non-negative square of their masses [17, 2]. This fact for the Minkowskian space ought to be checked in any particular field theory model, but from the estimation theory it is evident that:

$$\sigma^2(\hat{\theta}_n) \geq 0 , \quad (25)$$

which is always the case for *causal processes*.

Remark 1. *The index of the measurement channel:* The sample index n is the smallest index of the information channel in which the measurement is performed. It indicates that if the sample index might be additionally indexed, e.g. by the space-time index ν , then it would be impossible to perform the measurement in one of the subchannels ν assigned in that manner (without performing it simultaneously in the remaining subchannels which also possess the sample index n). The channel which is inseparable from the experimental point of view will be referred to as the *measurement channel*.

Remark 2. In case of the restriction of the analysis to only a part of the measurement channel, one should ascertain that the value of the Stam's information which is the subject of the analysis is positive. For example, in the case of the neglect of the temporally indexed part of the space-time measurement channel,

⁸ Thus the channel information capacity has the form:

$$I = \sum_{n=1}^N I_{Fn} = \sum_{n=1}^N \int_{\mathcal{B}} dy P(y|\Theta) \sum_{\nu=0}^3 \mathfrak{F}_{nn\nu} = \int_{\mathcal{B}} dy P(y|\Theta) \sum_{n=1}^N \mathfrak{F}_{nn} = \int_{\mathcal{B}} dy i , \quad (23)$$

where $i := P(\Theta) \sum_{n=1}^N \mathfrak{F}_{nn}$ is the *channel information density* [5, 8, 9].

The channel information capacity I_{Fn} is invariant under the smooth invertible mapping $Y \rightarrow X$, where X is the new variable [16]. It is also invariant under the space and time reflections.

the Stam's inequality obtained for the spacial components has the form:

$$\begin{aligned}
0 &\leq I_S = \sum_{n=1}^N I_{S_n} \\
&\leq \sum_{n=1}^N \int d\vec{y} P(\vec{y}|\vec{\Theta}) \sum_{i=1}^3 \frac{\partial \ln P(\vec{y}|\vec{\Theta})}{\partial \theta_{ni}} \frac{\partial \ln P(\vec{y}|\vec{\Theta})}{\partial \theta_{ni}} =: \sum_{n=1}^N I_{F_n} = I, \quad (26)
\end{aligned}$$

where the vector denotation means that the analysis of both in the sample space and in the parameter space has been reduced to the spacial part of the random variables and parameters, respectively.

Remark 3. *The symmetry conformity of I_S and I :* The error of the estimation of the expected position four-vector in the n -th measurement channel, and thus $I_{S_n} = 1/\sigma^2(\hat{\theta}_n)$, must be independent of the coordinate system in the Minkowski (or Euclidean) space. Therefore, I_{S_n} as defined in (21) and (19) is invariant under the Lorentz transformations (i.e. boosts and rotations) for the metric tensor given by (13) or for metric tensor given by (14) under the Galilean ones.

As the measurements in the sample are independent, the channel information capacity I is invariant under the Lorentz transformation in the space with the Minkowskian metric, or under the Galilean transformation in the Euclidean space, only if every I_{F_n} is invariant. The conditions of the invariance of I_{S_n} and I_{F_n} converge if in the inequalities given by (20), the equalities are attained. More information on the invariance of the CRLB under the displacement, space reflection, rotation and affine transformation as well as unitary transformation can be found in [18].

Remark 4. *Minimization of I with respect to N :* Every n -th term, $n = 1, \dots, N$, in the sum (24) brings, as the degree of freedom for I , its analytical contribution. If only the added degrees of freedom do not have an effect on the already existing ones, then because every I_{F_n} is non-negative, the channel information capacity I has an increasing tendency with an increase of N . The minimization criterion of I with respect to N was used by Frieden and Soffer as an additional condition in the selection of the equation of motion for the field theory model or the generation equation in the statistical physics realm [2]. This means that values of N above the minimal allowable one lack a physical meaning for the particular application and leaving them in the theory makes it unnecessarily complex. Yet, sometimes part of the consecutive values of N are also necessary. Some examples were given in the Introduction.

4. The kinematic information in the Frieden-Soffer approach

The basic physical assumption of the Frieden-Soffer approach. Let the data $y \equiv (\mathbf{y}_n)_{n=1}^N$ be the realization of the sample for the position of the system where $\mathbf{y}_n \equiv (\mathbf{y}_n^\nu)_{\nu=0}^3 \in \mathcal{Y} \equiv \mathbb{R}^4$. In accordance with the assumption proposed by Frieden and Soffer [2], their collection is carried out by the system alone

in accordance with the probability density distributions, $p_n(\mathbf{y}_n|\theta_n)$, $n = 1, \dots, N$. The content of the above assumption can be expressed as follows: *The system samples the space-time which is accessible to it “collecting the data and performing the statistical analysis”, in accordance with some information principles* (see [2, 7, 5, 8, 9]).

Note: The estimation is performed by a human and the EPI method should be perceived as only a *type of statistical analysis*.

The statistical procedure in which we are interested concerns the inference about $p_n(\mathbf{y}_n|\theta_n)$ on the basis of the data y using the likelihood function $P(y|\Theta)$. Therefore, using (4), we can record the derivatives of $\ln P$ standing in I_{Fn} in (18) as follows:

$$\frac{\partial \ln P(y|\Theta)}{\partial \theta_{n\nu}} = \frac{\partial}{\partial \theta_{n\nu}} \sum_{n=1}^N \ln p_n(\mathbf{y}_n|\theta_n) = \sum_{n=1}^N \frac{1}{p_n(\mathbf{y}_n|\theta_n)} \frac{\partial p_n(\mathbf{y}_n|\theta_n)}{\partial \theta_{n\nu}} \quad (27)$$

and applying the normalization $\int_{\mathcal{Y}} d^4\mathbf{y}_n p_n(\mathbf{y}_n|\theta_n) = 1$ of each of the marginal distributions we obtain the following form of the channel information capacity:

$$I = \sum_{n=1}^N \int_{\mathcal{Y}} d^4\mathbf{y}_n \frac{1}{p_n(\mathbf{y}_n|\theta_n)} \sum_{\nu, \mu=0}^3 \eta^{\nu\mu} \left(\frac{\partial p_n(\mathbf{y}_n|\theta_n)}{\partial \theta_n^\nu} \frac{\partial p_n(\mathbf{y}_n|\theta_n)}{\partial \theta_n^\mu} \right). \quad (28)$$

Finally, as the amplitudes $q_n(\mathbf{y}_n|\theta_n)$ of the measurement data $\mathbf{y}_n \in \mathcal{Y}$ are determined as in [2, 1, 3]:

$$p_n(\mathbf{y}_n|\theta_n) =: q_n^2(\mathbf{y}_n|\theta_n), \quad (29)$$

simple calculations give:

$$I = 4 \sum_{n=1}^N \int_{\mathcal{Y}} d^4\mathbf{y}_n \sum_{\nu, \mu=0}^3 \eta^{\nu\mu} \left(\frac{\partial q_n(\mathbf{y}_n|\theta_n)}{\partial \theta_n^\nu} \frac{\partial q_n(\mathbf{y}_n|\theta_n)}{\partial \theta_n^\mu} \right), \quad (30)$$

which is almost the key form of the channel information capacity of the Frieden-Soffer EPI method. It becomes obvious that by the construction the rank of the field of the system, which is the number of the amplitudes $(q_n(\mathbf{y}_n|\theta_n))_{n=1}^N$, is equal to the dimension N of the sample.

4.1. The kinematic form of the Fisher information

The estimation of physical equations of motion using the EPI method [2, 5, 8, 9] is often connected with the necessity of rewriting I , originally defined on the statistical space \mathcal{S} , in a form which uses the displacements defined on the base space \mathcal{B} of the sample. The task is performed as follows: Let $\mathbf{x}_n \equiv (\mathbf{x}_n^\nu)$ be the displacements (e.g. the additive fluctuations) of the data $\mathbf{y}_n \equiv (\mathbf{y}_n^\nu)$ from their expectation values θ_n^ν , i.e.:

$$\mathbf{y}_n^\nu = \theta_n^\nu + \mathbf{x}_n^\nu. \quad (31)$$

In accordance with (15), the displacements \mathbf{x}_n^ν are the Fisherian variables. Appealing to the “chain rule” for the derivative [2]:

$$\frac{\partial}{\partial \theta_n^\nu} = \frac{\partial(\mathbf{y}_n^\nu - \theta_n^\nu)}{\partial \theta_n^\nu} \frac{\partial}{\partial(\mathbf{y}_n^\nu - \theta_n^\nu)} = - \frac{\partial}{\partial(\mathbf{y}_n^\nu - \theta_n^\nu)} = - \frac{\partial}{\partial \mathbf{x}_n^\nu}, \quad (32)$$

and taking into account $d^4 \mathbf{x}_n = d^4 \mathbf{y}_n$, we can switch from the statistical form (30) to the *kinematic form* of the channel information capacity:

$$I = 4 \sum_{n=1}^N \int_{\mathcal{X}_n} d^4 \mathbf{x}_n \sum_{\nu, \mu=0}^3 \eta^{\nu\mu} \frac{\partial q_n(\mathbf{x}_n + \theta_n | \theta_n)}{\partial \mathbf{x}_n^\nu} \frac{\partial q_n(\mathbf{x}_n + \theta_n | \theta_n)}{\partial \mathbf{x}_n^\mu}, \quad (33)$$

where \mathcal{X}_n is the space of the \mathbf{x}_n displacements. In rewriting (30) in the form of (33), the equality:

$$q_n(\mathbf{y}_n | \theta_n) = q_n(\mathbf{x}_n + \theta_n | \theta_n) \quad (34)$$

has been used.

Assuming that the range of the variability of all \mathbf{x}_n^ν is the same for every n , we can disregard the index n for these variables (but not for the amplitudes q_n) and thus obtain the formula:

$$I = 4 \sum_{n=1}^N \int_{\mathcal{X}} d^4 \mathbf{x} \sum_{\nu, \mu=0}^3 \eta^{\nu\mu} \frac{\partial q_n(\mathbf{x} + \theta_n | \theta_n)}{\partial \mathbf{x}^\nu} \frac{\partial q_n(\mathbf{x} + \theta_n | \theta_n)}{\partial \mathbf{x}^\mu}, \quad (35)$$

where \mathcal{X} is the space of displacements \mathbf{x} . The resulting formula (35) indicates that the Fisher-Rao metric (8) on the statistical space \mathcal{S} generates the *kinetic energy metric*. Let us now note that the kinetic term for every particular amplitude q_n enters with different θ_n , in principle. Hence, in fact N kinetic terms have been obtained, one for every amplitude $q_n(\mathbf{x} + \theta_n | \theta_n)$.

The EPI model from which, e.g. a particular field theory model appears, is usually⁹ built over the displacements space \mathcal{X} , which in our case is the Minkowski space-time R^4 . Although the statistical model \mathcal{S} is transformed from being defined on the parameter space $V_\Theta \equiv \mathfrak{R}^4$ to the space of displacements $\mathcal{X} \equiv \mathfrak{R}^4$, the original sample space \mathcal{B} remains its base space both before and after this redefinition.

In this way, the basic tool for the EPI model estimation connected with the derivation of both the generating equations of statistical physics [2, 11] and the equations of motion for field theory models [2, 8, 10], e.g. the Maxwell electrodynamics [2], is obtained.

A simplifying notation will now be introduced:

$$q_n(\mathbf{x}) \equiv q_{\theta_n}(\mathbf{x}) \equiv q_n(\mathbf{x} + \theta_n | \theta_n), \quad (36)$$

⁹The analysis of the EPR-Bohm experiment takes place on a parameter space V_Θ [2, 8].

which leaves the whole information on θ_n that characterizes $q_n(\mathbf{x} + \theta_n|\theta_n)$ in the index n of the amplitude $q_n(\mathbf{x})$ (and similarly for the original distribution $p_n(\mathbf{x})$). With this simplifying notation, the formula (35) can be rewritten as follows:

$$I = 4 \sum_{n=1}^N \int_{\mathcal{X}} d^4\mathbf{x} \sum_{\nu, \mu=0}^3 \eta^{\nu\mu} \frac{\partial q_n(\mathbf{x})}{\partial \mathbf{x}^\nu} \frac{\partial q_n(\mathbf{x})}{\partial \mathbf{x}^\mu}, \quad (37)$$

which is understood in accordance with (35).

Note on the shift invariance: In the above derivation, the assumption used in [2] on the *invariance of the distribution under the shift (displacement)*: $p_n(\mathbf{x}_n) \equiv p_{x_n}(\mathbf{x}_n|\theta_n) = p_n(\mathbf{x}_n + \theta_n|\theta_n) = p_n(\mathbf{y}_n|\theta_n)$, where $\mathbf{x}_n^\nu \equiv \mathbf{y}_n^\nu - \theta_n^\nu$, has not been used.

4.2. The probability form of the kinematic channel information capacity

Here, the starting point is the form of I given by (28). We have to move on to the additive displacements $\mathbf{x}_n \equiv (\mathbf{x}_n^\nu)$ (31) and use the “chain rule” (32). As was mentioned previously, assuming that the range of the variability of all \mathbf{x}_n^ν is the same for every n and disregarding the variables index n , but leaving the information about θ_n in the subscript of the point distributions p_n , we obtain the *kinematic form* of the channel information capacity expressed as the functional of the probabilities:

$$I = \sum_{n=1}^N \int_{\mathcal{X}} d^4\mathbf{x} \frac{1}{p_n(\mathbf{x})} \sum_{\nu, \mu=0}^3 \eta^{\nu\mu} \left(\frac{\partial p_n(\mathbf{x})}{\partial \mathbf{x}^\nu} \frac{\partial p_n(\mathbf{x})}{\partial \mathbf{x}^\mu} \right), \quad (38)$$

where the simplifying notation $p_n(\mathbf{x}) \equiv p_{\theta_n}(\mathbf{x}) \equiv p_n(\mathbf{x} + \theta_n|\theta_n)$, with the same meaning as (36) in (37) has been used. The above form is used as the primary one in the construction of I for Maxwell electrodynamics (see below) and in the weak field limit of the gravitation theory [8].

4.3. The channel information capacity for Maxwell electrodynamics

The Frieden-Soffer EPI method of the derivation of the Maxwell equations of electrodynamics was presented in [2]. However, in [2] the form of I is the Euclidean one. What was lacking for the construction of the Minkowski form of I is the notion of the measurement channel presented in previous sections from which the fully relativistic covariant form of the channel information capacity required for the Maxwell model follows [8]. Below, we present its construction from which the EPI method for the Maxwell equations gains its full Fisherian statistical validity.

We begin with the channel information capacity written in the basic form given by (38). The proof that in order to obtain the Maxwell equations of motion, the field of the rank equal to $N = 4$ with real amplitudes q_n , $n = 1, 2, 3, 4$

has to be analyzed, is given in [2], where one of the assumptions is that the gauge fields are proportional to these amplitudes:

$$q_\nu(\mathbf{x}) = a A_\nu(\mathbf{x}) , \quad \text{where } \nu \equiv n - 1 = 0, 1, 2, 3 , \quad (39)$$

where a is a constant and the Minkowski sample index ν is introduced. Using the Minkowski metric $(\eta^{\nu\mu})$, we now define the amplitudes $q^\nu(\mathbf{x})$ which are *dual* to $q_\nu(\mathbf{x})$:

$$\begin{aligned} q^\nu(\mathbf{x}) &\equiv \sum_{\mu=0}^3 \eta^{\nu\mu} q_\mu(\mathbf{x}) = a \sum_{\mu=0}^3 \eta^{\nu\mu} A_\mu(\mathbf{x}) \\ &\equiv a A^\nu(\mathbf{x}) , \quad \text{where } \nu \equiv n - 1 = 0, 1, 2, 3 , \end{aligned} \quad (40)$$

where the dual gauge fields $A^\mu(\mathbf{x})$ have been introduced:

$$A^\nu(\mathbf{x}) \equiv \sum_{\mu=0}^3 \eta^{\nu\mu} A_\mu(\mathbf{x}) , \quad \text{where } \nu = 0, 1, 2, 3 . \quad (41)$$

The amplitudes $q_\nu(\mathbf{x})$ are connected with the point probability distributions $p_n(\mathbf{x})$ in the following way:

$$\begin{aligned} p_n(\mathbf{x}) &\equiv p_{q_\nu}(\mathbf{x}) = q_\nu(\mathbf{x}) q_\nu(\mathbf{x}) \\ &= a^2 A_\nu(\mathbf{x}) A_\nu(\mathbf{x}) , \quad \text{where } \nu \equiv n - 1 = 0, 1, 2, 3 . \end{aligned} \quad (42)$$

Consequently, we observe that in the EPI method applied to Maxwell electrodynamics, the sample index n becomes the space-time one. Thus, the form of the channel information capacity has to take into account the additional estimation in the space-time channels taking, in accordance with the general prescriptions of Section 3.1, the covariant form:

$$\begin{aligned} I &= \sum_{\mu=0}^3 \int_{\mathcal{X}} d^4\mathbf{x} \eta^{\mu\mu} \frac{1}{p_{q_\mu}(\mathbf{x})} \sum_{\nu=0}^3 \left(\frac{\partial p_{q_\mu}(\mathbf{x})}{\partial \mathbf{x}_\nu} \frac{\partial p_{q_\mu}(\mathbf{x})}{\partial \mathbf{x}^\nu} \right) \\ &= 4 \sum_{\mu=0}^3 \int_{\mathcal{X}} d^4\mathbf{x} \sum_{\nu=0}^3 \left(\frac{\partial q_\mu(\mathbf{x})}{\partial \mathbf{x}_\nu} \frac{\partial q^\mu(\mathbf{x})}{\partial \mathbf{x}^\nu} \right) . \end{aligned} \quad (43)$$

Finally, in accordance with (43) and (40), the channel information capacity I is as follows:

$$I = 4 a^2 \sum_{\mu=0}^3 \int_{\mathcal{X}} d^4\mathbf{x} \sum_{\nu=0}^3 \left(\frac{\partial A_\mu(\mathbf{x})}{\partial \mathbf{x}_\nu} \frac{\partial A^\mu(\mathbf{x})}{\partial \mathbf{x}^\nu} \right) . \quad (44)$$

It must be stressed that the proportionality $q_\nu(\mathbf{x}) = a A_\nu(\mathbf{x})$ and the normalization condition:

$$(1/4) \sum_{\nu=0}^3 \int_{\mathcal{X}} d^4\mathbf{x} q_\nu^2(\mathbf{x}) = \sum_{\nu=0}^3 \int_{\mathcal{X}} d^4\mathbf{x} A_\nu^2(\mathbf{x}) = 1 , \quad (45)$$

where $a = 2$, pose a question on the meaning of the localization of the photon and the existence of its wave function. These points have recently been discussed in the literature [19]. For example, the discussion in [19] supports the view that places the Maxwell equations on the same footing as the Dirac one; a fact previously noted by Sakurai [15]. It is worth noting that the normalization condition¹⁰ (45) of the four-potential A_ν harmonizes the value of the proportionality constant $a = 2$ with the rank $N = 4$ of the light field whose value of a also occurred previously in [2] but as the result of harmonizing the EPI result¹¹ with the Maxwell equations.

Now, both the variational and structural (see Introduction and [2, 8, 10]) information principles have to be self-consistently solved with the Lorentz condition [2]:

$$\sum_{\nu, \mu=0}^3 \eta^{\mu\nu} \partial_\mu A_\nu = 0 \quad (46)$$

additionally imposed. This task, with the proper form of the structural information for the Maxwell equations additionally found, is presented in [2]. In that work, the Maxwell equations were obtained by solving the EPI information principles and a discussion of the solutions for the gauge fields can be also found there. What has been lacking is the construction of the Minkowski form of the channel information capacity, which was presented above.

5. The Fourier information

Let us now consider the particle as a system described by the field of the rank N , which is the set¹² of the amplitudes $q_n(\mathbf{x})$, $n = 1, 2, \dots, N$ determined in the position space-time \mathcal{X} of displacements $\mathbf{x} \equiv (\mathbf{x}^\mu)_{\mu=0}^3 = (ct, \mathbf{x}^1, \mathbf{x}^2, \mathbf{x}^3)$ as in Section 4.1 that possesses the channel information capacity (37). The complex Fourier transforms $\tilde{q}_n(\mathbf{p})$ of the real functions $q_n(\mathbf{x})$, where $\mathbf{p} \equiv (\wp^\mu)_{\mu=0}^3 = (\frac{E}{c}, \wp^1, \wp^2, \wp^3)$ is the four-momentum which belongs to the energy-momentum space \mathcal{P} conjugated to the position space-time \mathcal{X} , have the form:

$$\tilde{q}_n(\mathbf{p}) = \frac{1}{(2\pi\hbar)^2} \int_{\mathcal{X}} d^4\mathbf{x} q_n(\mathbf{x}) e^{i(\sum_{\nu=0}^3 \mathbf{x}^\nu \wp_\nu)/\hbar}, \quad (47)$$

where $\sum_{\nu=0}^3 \mathbf{x}^\nu \wp_\nu = Et - \sum_{l=1}^3 \mathbf{x}^l \wp^l$ and \hbar is the Planck constant. The Fourier transform is the *unitary transformation which preserves the measure*

¹⁰The normalization (45) of the four-potential A_ν to unity might not necessarily occur. The indispensable condition for $q_\nu(\mathbf{x})$ is that the necessary means can be calculated [20].

¹¹Which is the result of solving [2] the structural and variational information principles [2, 5, 9] together with $\partial_\mu A^\mu = 0$, which is the Lorentz one.

¹²For the case of complex wave functions see [2].

on the L^2 space of functions integrable with the square, i.e.:

$$\int_{\mathcal{X}} d^4\mathbf{x} q_n^*(\mathbf{x}) q_m(\mathbf{x}) = \int_{\mathcal{P}} d^4\mathbf{p} \tilde{q}_n^*(\mathbf{p}) \tilde{q}_m(\mathbf{p}) . \quad (48)$$

Hence, applying the condition of the probability normalization¹³:

$$\frac{1}{N} \sum_{n=1}^N \int_{\mathcal{X}} d^4\mathbf{x} q_n^2(\mathbf{x}) = 1 , \quad (49)$$

we obtain (as the consequence of the Parseval's theorem):

$$\frac{1}{N} \sum_{n=1}^N \int_{\mathcal{X}} d^4\mathbf{x} q_n^2(\mathbf{x}) = \frac{1}{N} \sum_{n=1}^N \int_{\mathcal{P}} d^4\mathbf{p} |\tilde{q}_n(\mathbf{p})|^2 = 1 , \quad (50)$$

where $|q_n(\mathbf{x})|^2 \equiv q_n^2(\mathbf{x})$ and $|\tilde{q}_n(\mathbf{p})|^2 \equiv \tilde{q}_n^*(\mathbf{p}) \tilde{q}_n(\mathbf{p})$.

Using (47) we can now record I given by (37) as follows:

$$I[q(\mathbf{x})] = I[\tilde{q}(\mathbf{p})] = \frac{4}{\hbar^2} \int_{\mathcal{P}} d^4\mathbf{p} \sum_{n=1}^N |\tilde{q}_n(\mathbf{p})|^2 \left(\frac{E^2}{c^2} - \vec{\wp}^2 \right) , \quad (51)$$

where $\vec{\wp}^2 = \sum_{k=1}^3 \wp_k \wp^k$.

The determination of the square of the particle mass: As the channel information capacity I is from the definition the sum of the expected values (see (18) and (24)), the squared *mass* of a particle, defined as [2]:

$$m^2 := \frac{1}{N c^2} \int_{\mathcal{P}} d^4\mathbf{p} \sum_{n=1}^N |\tilde{q}_n(\mathbf{p})|^2 \left(\frac{E^2}{c^2} - \vec{\wp}^2 \right) \quad (52)$$

is constant and does not depend on the statistical fluctuations of the energy E and momentum $\vec{\wp}$, i.e. at least as the mean after the integration is performed. Thus, for a free particle we can record (51) as follows:

$$I[q(\mathbf{x})] = I[\tilde{q}(\mathbf{p})] = 4N \left(\frac{m c}{\hbar} \right)^2 = \text{const} . \quad (53)$$

¹³ **Note on the displacement probability distribution in the system:** Appealing to the theorem of the total probability, the probability *density* distribution of the displacement (or fluctuation) in the system can be written as follows [2]:

$$p(\mathbf{x}) = \sum_{n=1}^N p(\mathbf{x}|\theta_n) r(\theta_n) = \sum_{n=1}^N p_n(\mathbf{x}_n|\theta_n) r(\theta_n) = \frac{1}{N} \sum_{n=1}^N q_n^2(\mathbf{x}_n|\theta_n) = \frac{1}{N} \sum_{n=1}^N q_n^2(\mathbf{x}) .$$

Function $r(\theta_n) = \frac{1}{N}$ can be referred to as the “ignorance” function as its form is a reflection of the total lack of knowledge on which out from N possible values of θ_n appears in a specific n -th experiment in an N -dimensional sample.

Therefore, we observe that the causality relation (25) entails $I \geq 0$ in (22) and then, in accord with (53) the condition $m^2 \geq 0$, i.e. the lack of tachions in the theory [17]. It must be pointed out that, according to (52), the zeroing of particle mass would be impossible for the Euclidean space-time metric (14).

The Fourier information with q : The condition (53) means that:

$$K_F \equiv I[q(\mathbf{x})] - I[\tilde{q}(\mathbf{p})] = 0 , \quad (54)$$

what, using the constancy of $4N(\frac{mc}{\hbar})^2$ and (50) can be rewritten as the condition fulfilled by the free field of the rank N :

$$K_F = \int_{\mathcal{X}} d^4\mathbf{x} k_F = 0 . \quad (55)$$

Here, the quantity K_F defines the so-called *Fourier information* (F), where its density k_F is equal to:

$$k_F = 4 \sum_{n=1}^N \left[\sum_{\nu=0}^3 \frac{\partial q_n(\mathbf{x})}{\partial \mathbf{x}_\nu} \frac{\partial q_n(\mathbf{x})}{\partial \mathbf{x}^\nu} - \left(\frac{mc}{\hbar}\right)^2 q_n^2(\mathbf{x}) \right] . \quad (56)$$

Remark 5. Ignoring the fact that from the above calculations m^2 emerges as the mean (52), the equation (54), and consequently (55), is the reflection of the Parseval's theorem and as such it is a tautology. This means that the Fourier transformation reflects the change of the basis in the statistical space \mathcal{S} only. Therefore, by itself only the condition (55) does not superimpose any new constraint *unless* $4N(\frac{mc}{\hbar})^2$ is additionally determined as the structural information of the system, which in this particular case defines the type of the field, the scalar one [2, 7, 5, 8, 9].

6. Conclusions

A system without a structure dissolves itself, hence its equation of motion requires besides the kinematical term (which is the channel information capacity I) a structural one. Intuitively, "during putting the structure upon" the information on the system has to be maximized and this is performed by the addition of the structural information Q , which acts under the condition of zeroing of the *observed structural information principle* [7, 5, 8, 9]. Both the analysis of the analyticity of the log-likelihood function [5] and the Rao-Fisher metricity of the statistical (sub)space \mathcal{S} [1] allow for the formal construction of this observed differential form of the structural information principle [5, 9], which, when lifted to the expected level, connects the channel information capacity I with the structural information Q . However, when this structural differential constraint is established then from all distributions in \mathcal{S} the one which minimizes total physical information $K \equiv I + Q$ is chosen. It is achieved via the *variational principle* $I + Q \rightarrow \min$.

The above two information principles, i.e. firstly, the observed and expected structural information principle and secondly, the variational one, form the core of the EPI method of estimation [5, 8, 9] (see Introduction). In order to estimate correctly the distribution of the system, the EPI method uses these principles, estimating the equation of motions for the particular field theory model, or the equation generating the distribution in the particular model of the statistical physics. The whole procedure depends greatly on the form of the channel information capacity I used in the particular physical model, which has to take into account the geometrical and physical preconditions, among these the metric of the base space \mathcal{Y} and the (internal) symmetries of the physical system. For example, different representations of the Lorentz transformation, which is the isometry of I , lead to its specific decomposition giving Klein-Gordon or Dirac equation [2, 8, 10]. Nonetheless, we cannot rule out the possibility that the opposite is also indisputable. For instance, from the basic property of the non-negativity of I , (22), preceded by the causality property (25), which is the statistical requirement of any successful observation, there follows the very important physical property (53) of the non-existence of tachion in the field theory models.

In accord with *Remark 5* underneath (56), only when the structural information principle [5, 8, 9] fulfils also the mass relation (52) which is put as the constrain upon the Fourier transformation (47), the entanglement¹⁴ among the momentum degrees of freedom in the system appears.

Now, let us discuss a massless particle, e.g. photon. If, in agreement with Section 4.3, the condition (55)-(56) is adopted for a particle with a mass $m = 0$ and the amplitude is interpreted according to (39) and (45) as characterizing the photon, it would then necessitate answers to both the question on the nature of the photon wave function [19] and the signature of the space-time metric from experiments. Indeed, in the Minkowski space-time metric (13), according to (52), the only possibility for a particle to be massless is that it fulfills the condition $E^2/c^2 - \vec{p}^2 = 0$ for all its Fourier monochromatic modes, if only the Fourier modes of massless particles are to possess a physical interpretation [8]. However, this condition means that the Fourier modes are not entangled (in contrast to the massive particle case) and in principle, the possibility to detect every individual mode should exist. If the individual Fourier mode frequencies of a light pulse were not detected, it would have meant that they are not physical objects [19]. This would lead to a serious problem for the quantum interpretation of a photon as the physical realization of a particular Fourier

¹⁴It follows from Section 5 that the Fourier transformation forms a kind of self-entanglement (between two representations of I , one of positions and the other of momentums) of the realized values of the variables of the system [2] and leads to (51). Nevertheless, without putting the structural information principle on the system, which is the analyticity requirement of the log-likelihood function and without the Rao-Fisher metricity of \mathcal{S} , the relation (51) remains a tautological sentence only. Indeed, in the case of the relation discussed in (51), the Fourier transformation has to relate the Fisher information matrix, which is in the second order term of the Taylor expansion of the log-likelihood function [5] with remaining parts of this expansion.

mode. This problem recently occurred in the light beam experiments discussed in [19], which suggests that the Fourier decomposed frequencies of a pulse do not represent actual optical frequencies, possibly implying that a real photon is a "lump of electromagnetic substance" without Fourier decomposition [19, 21]. Finally, in general, the channel information capacity and the structural information form the basic concepts in the analysis of entanglement in the system. Nonetheless, e.g. for analyses of this phenomenon for the fermionic particle, the complex amplitude should sometimes be introduced [2, 8].

Acknowledgments

The contribution of J. Syska to the work has been supported by L. J. Ch. The work of Jan Śladkowski has been supported by the project **Quantum games: theory and implementations** financed by the **National Science Center** under the contract no **DEC-2011/01/B/ST6/07197**, which has supported also J. Syska in part.

References

- [1] S. Amari, H. Nagaoka, *Methods of information geometry*, translations of Mathematical monographs, Vol.191, Oxford Univ.Press, 2000. A. Fujiwara and H. Nagaoka, *Phys.Lett.A* 201, 1995, 119-124. A. Fujiwara and H. Nagaoka, *J.Math.Phys.* 40, 1999, 4227-4239. K. Matsumoto, *A geometrical approach to quantum estimation theory*, PhD thesis, Univ. of Tokyo, 1998.
- [2] B.R. Frieden, *Found.Phys.*, Vol.16, No.9, 1986, 883-903. B.R. Frieden, *Phys.Rev.A* 41, 1990, 4265-4276. B.R. Frieden, B.H. Soffer, *Phys.Rev.E* 52, 1995, 2274-2286. B.R. Frieden, *Phys.Rev.A* 66, 2002, p.022107. B.R. Frieden, A. Plastino, A.R. Plastino and B.H. Soffer, *Phys.Rev. E* 66, 2002, p.046128. B.R. Frieden, A. Plastino, A.R. Plastino and B.H. Soffer, *Phys.Lett.A* 304, 2002, 73-78. B.R. Frieden, *Science from Fisher information: A unification*, Cambridge Univ.Press, 2004.
- [3] I. Bengtsson, K. Życzkowski, *Geometry of quantum states*, Cambridge Univ.Press, 2006.
- [4] D. M. Endres and J. E. Schindelin, *IEEE Trans.Inf.Theory* 49, 2003, p.1858.
- [5] E.W. Piotrowski, J. Śladkowski, J. Syska, S. Zając, *Phys.Stat.Sol.(b)*, 246, No.5, doi:10.1002/pssb.200881566, 2009, 1033-1037, arXiv:0811.3554 [physics.data-an].
- [6] E. Jaynes, *Phys.Rev.* 106, 1957, 620-630. E. Jaynes, *Phys.Rev.* 108, 1957, 171-190.
- [7] J. Syska, *Phys.Stat.Sol.(b)*, 244, No.7, doi:10.1002/pssb.200674646, 2007, 2531-2537, arXiv:0811.3600 [physics.data-an].

- [8] J. Syska, Maximum likelihood method and Fisher's information in physics and econophysics (*polish version*), University of Silesia, <http://el.us.edu.pl/ekonofizyka/images/f/f2/Fisher.pdf>, 2011.
- [9] J. Śladrkowski, J. Syska, Information principles in the Extreme Physical Information method, *in preparation*. Id. [5] and Id. [8].
- [10] J. Syska, R. Szafron, Informational criteria for the field theory models building, *in preparation*.
- [11] I. Bednarek, E.W. Piotrowski, J. Śladrkowski, J. Syska, Generalization of the Aoki-Yoshikawa Sectoral Productivity Model, *in preparation*. J. Syska in [8] and J. Śladrkowski in International Conf. on Statistical Physics SigmaPhi, 2011.
- [12] E.W. Piotrowski, J. Śladrkowski, Acta Phys.Pol.B, Vol.32, No.2, 2001, 597-604. H. Kleinert, X.J. Chen, Boltzmann distribution and market temperature, 2007, arXiv:physics/0609209v2 [physics.soc-ph].
E.W. Piotrowski, J. Śladrkowski, J. Syska, Physica A 318, 2003, 516-528. E.W. Piotrowski, J. Śladrkowski, J. Syska, Physica A, no.389, doi:10.1016/j.physa.2010.06.062, 2010, 4904-4912.
- [13] Y. Pawitan, In all likelihood: Statistical modelling and inference using likelihood, Oxford Univ.Press, 2001.
- [14] A.J. Stam, Some mathematical properties of quantities of information, Ph.D. dissertation, Technical University of Delft, 1959. A.J. Stam, Inform. and Control 2, 1959, 101–112.
- [15] J.J. Sakurai, Advanced quantum mechanics, Addison-Wesley Publishing Company, 1967, p.169.
- [16] R.F. Streater, Statistical dynamics. A stochastic approach to nonequilibrium thermodynamics, 2nd ed., Imperial College Press, 2009.
- [17] J. Syska, Int.J.Theor.Phys., Vol. 49, Issue 9, doi:10.1007/s10773-010-0400-8, Open Access, 2010, 2131-2157.
- [18] F. Pennini, A. Plastino, B.H. Soffer, C. Vignat, Phys.Lett.A 373, 2009, 817-820.
- [19] The nature of light. What is a photon?, Ch. Roychoudhuri, A.F. Kracklauer, K. Creath, ed., CRC Press, Taylor&Francis Group, 2008, 363-377.
- [20] U. Leonhardt, Essential quantum optics. From quantum measurements to black holes, Cambridge Univ.Press., 2010.
- [21] J. Syska, Frontiers in field theory, ed. O. Kovras, Nova Science Publishers, Nowy Jork, chapter 6, 2005, 125-154, and Focus on Boson Research, Nova Publishers, 2006, 233-258, arXiv:hep-ph/0205313.